DOCUMENT RESUME

ED 285 153 CS 210 485

AUTHOR Fuller, Deborah Ann
TITLE Evaluation and Curriculum: A Case Study in Primary Trait Scoring.
PUB DATE Jan 87
NOTE 39p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Case Studies; Curriculum Evaluation; Elementary Education; Evaluation Criteria; *Evaluation Methods; Grade 7; Group Discussion; *Holistic Evaluation; Instructional Improvement; *Reader Response; Reader Text Relationship; Student Writing Models; Teacher Attitudes; Test Reliability; Test Validity; *Writing Evaluation; Writing Research
IDENTIFIERS *Primary Trait Scoring; Writing Attitudes

ABSTRACT

A case study examined the effects of primary trait scoring on evaluating seventh grade students' compositions. Primary trait scoring as used in the Metropolitan School District of Washington Township in Indianapolis is innovative in that (1) the scoring guide is assignment specific, written after adult raters read and discussed many papers; (2) the same scorers have been involved in the procedure for many years; (3) raters are encouraged to act as real readers and to discuss papers as meaningful texts, and (4) scorers are required to try to ignore mechanics and other secondary writing features. Eighty papers, originally scored in a pilot study conducted from 1980 through 1982, were chosen for rescoring in 1984. Raters were closely monitored, discussions held during scoring sessions were analyzed, and the entire scoring process was evaluated. Results indicated that primary trait scoring was a reliable method. Over half of the papers received scores identical to their previous ratings. The evaluation process caused the participants to rethink beliefs about writing and writing instruction and to examine and refine those beliefs through group discussions. Findings also showed that because the method validated the use of whole texts and because participants acted as leaders for change within a school, more teachers were involved in having students write whole compositions, and curricular and methodological change occurred in the classroom. (References, writing samples, and the scoring guide are appended.) (NKA)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

ED285153

Evaluation and Curriculum: A Case Study in Primary Trait Scoring

by Deborah Ann Fuller

Metropolitan School District
of Washington Township
3801 East 79th Street
Indianapolis, IN 46240

January, 1987

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Deborah Ann Fuller

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

CS 210 485

Any method used by schools to evaluate students' writing will provide guidelines for the evaluators' thinking. In the most static case, the reporting of subtest scores from a standardized test, writing effectiveness is reduced in the evaluators' minds to spelling, usage, punctuation, and capitalization. Those who report writing effectiveness in terms of such subtest scores have little choice but to equate effective writing with the use of standard spelling, usage, punctuation, and capitalization. The method of evaluating the product has influenced the evaluators' perceptions of effective student writing.

To overcome such limited notions of effective writing, school districts might use some form of holistic scoring. By having the evaluators read and react to whole texts, the school administrators assume that the score of the paper judged as a whole reflects more accurately a student's writing ability than scores that would result from an analysis of the essay's parts. Those holistic scores, however, may reflect more than the skills of the writers. Since the evaluators are readers of those papers, their reading processes will affect their decisions as evaluators. An understanding of reading, then, is necessary for a more complete understanding of the resulting scores.

Louise Rosenblatt (1978) in her book The Reader, The Text, The Poem used the term "transaction" to describe the reading process. When a reader and a text come together, the result is a new event: the poem. With this transactional view of reading, it is clear that different readers would have different "poems," and that even the same reader on different occasions would develop different "poems." According to Rosenblatt, the reader does not remain untouched by the experience of reading the text. One cannot go back and undo the reading experience, nor can one prevent one reading experience from affecting another. Harste and Stephens (1984) argue that "what we bring to the setting affects what we perceive, and what we perceive affects what we know." Teachers who read student compositions bring many things to that task: their experiences reading other student compositions, a view of their role as teachers, their knowledge about writing

gleaned from textbook lessons, and their own experiences as writers. All of these will affect the readers' perceptions of the student compositions, the readers' creations of the new "poems."

Using this transactional description of reading, one can see that the evaluators involved in any form of holistic scoring will have their reading affected not only by the scoring guide, but also by their previous experiences. If, as part of the scoring procedure, the readers are involved in discussing certain papers and debating their effectiveness, the resulting product scores reflect not only the success of the writer, but also the growth or change that has taken place in the readers. Test administrators can only guess at the subsequent effect of that new understanding on the evaluation of later papers.

Dealing with only the scores from a writing measure also masks the impact that the evaluation procedure has on the instructional process. If the guidelines set to evaluate a written product influence the evaluators' perceptions of an effective paper, then that new understanding may influence instruction. If teachers are involved in the evaluation, either as the recipient of the final scores or as evaluators, those teachers may direct their teaching to what has been deemed important. Burke and Hartste (1977) argue that both learning and teaching are theoretically based. Teachers base their methods, consciously or unconsciously, on the theoretical models they hold concerning how students learn language. Students bring their own set of expectations to the learning situation. Involvement in a form of holistic scoring may shape teachers' theoretical models which, in turn, would affect the students' expectations. Product scores, such as those from holistically scored papers, may not simply report; they may influence and direct instruction. Without an analysis of how we measure the effectiveness of writing, including the changes occurring in the evaluators and in the program, we are attempting to evaluate the success of instruction through the artifacts of the writing process, unaware that this static picture may not only misrepresent the writing process, but may also affect that process.

The purpose of this study is to examine the effects of one method of evaluating compositions, a form of holistic scoring called primary trait scoring, as it is used to evaluate

seventh grade compositions in the Metropolitan School District of Washington Township in Indianapolis, Indiana.

The form of primary trait scoring used in this school district is unique for several reasons:

1) the scoring guide is assignment specific, having been written after the adults read and discussed many papers. The guide describes the student papers rather than gives the adults' beliefs prior to reading the papers of what an effective paper would be. 2) the same scorers have been involved over the years that the procedure has been in place. 3) the raters are encouraged to react as real readers and to discuss the papers as meaningful texts. 4) the scorers are forced to try to ignore mechanics, usage, spelling, and other secondary features with frequent discussions of problem papers acting as a safeguard to the process.

Primary trait scoring was chosen because it is purported to describe the effectiveness of a composition as it relates to the specific purpose and audience of the assignment, not the writer's control of spelling, usage, and punctuation in the first draft stage of the paper. To see if the evaluators were able to ignore these "secondary" traits, I looked at a sample of those compositions from two years and then from a rescoring of some of those papers to see if the scorers were able to block their concern for surface level conventions. A check of the rescored papers allowed me to check the reliability of the scores from one scoring session to the next. I also examined the transcript of a scoring session to map the interplay between the raters while they were using the scoring guide to evaluate the papers and to reach consensus about their own perceptions of the papers. Finally, I worked with the raters individually as they read sample papers and shared their perceptions of the papers. These individual interviews also gave me information concerning the impact of the primary trait scoring sessions on the readers as classroom teachers and the subsequent effect on writing instruction district-wide.

Procedures for a Process Evaluation of Primary Trait Scoring

In response to the demands for minimum competency testing, the middle school teachers and curriculum coordinators of the Metropolitan School District of Washington Township wanted to supplement the standardized testing program with its objective measure of editing skills. They wanted a composition measure that would examine the effectiveness of students' essay writing. Primary trait scoring, a form of holistic scoring, was chosen as the measure. In the spring of 1979, seventh grade students were given the following prompt:

> Write about something you either like or dislike intensely. It may not be a person. Describe it in such a way that the reader will feel the same way as you do.

The students' papers were given three-digit numbers for identification purposes so that the writers' names could be masked to insure the anonymity of the student-author during the scoring session. That summer, a group of English teachers read the papers and devised a descriptive scoring guide from those papers that could be used to reliably score papers from students who had written on the same assignment.

The finished scoring guide (See Appendix A) consists of more than a description of the papers in each category. Examples of papers in each category are also part of the guide so that an evaluator could compare papers in order to come to a decision regarding the rating of a paper.

In the springs of 1980, 1981, and 1982, all seventh grade students in the district were required to write on the same composition topic. The scoring guide from the pilot study was used to rate the papers. Of the original readers, six remained constant.

In order to examine the interrelationships among the evaluation procedure, the teacher/evaluators, and classroom instruction, I first looked at the products, sample papers from 1981 and 1982. Then five people involved in the scoring since the pilot study rescored some of the papers in 1984, and I was able to examine the processes involved.

Product Analysis

Initially I was interested in looking at the papers themselves so that they could be described in ways that would augment the description found in the scoring guide. It was important to verify that primary traits, not secondary traits, influenced the readers in order to demonstrate that this scoring method did what it purported to do. To be able to describe the papers in more depth, I chose ten at random from the 1981 and 1982 papers from each of the following ratings: one, two, three, and four. Papers that were assigned a zero were not included in this study since the zero category was reserved for those papers whose authors did not follow the assignment.

For these 80 papers, I examined the students' choices of topics, the beginning and ending sentences of those papers, as well as the following surface features: number of words, percentage of unconventional to conventional spellings, number and type of usage and punctuation errors, and syntactical complexity as measured by t-units. The papers were examined as groups according to the year in which they were written and according to the groups of papers discussed in the 1984 rescoring session. As a result of this grouping, I had information available for each score (one - four) concerning surface features for six groups of papers: the 1981 papers, the 1982 papers, the 1981 and 1982 papers combined, the ten papers rescored by group consensus in 1984, the 36 papers rescored using the usual procedures in 1984, and the combination of all papers rescored in 1984.

Process Analysis

With the help of five people who had been involved in the development of the scoring guide in 1979 and who had worked with the scoring sessions in the subsequent summers, an afternoon session was arranged during which papers drawn from the 80 that I had analyzed would be rescored. The procedure was as follows:

a) All participants received ten papers to score using the scoring guide. The ten papers were chosen at random from the 80 that I had analyzed. The readers shared their scores for the

papers and discussed any discrepancies until they reached consensus on a score. They were not told what scores any of the papers had received originally.

b) After having become familiar again with the scoring guide through the scoring and discussion of the initial ten papers, each reader received a packet of six papers chosen at random from the remaining 70 papers. Regular scoring procedures were followed: after the initial reader had scored the packet, I kept those scores and gave the packet to a second reader. If there was a discrepancy between scores given for any paper by the first and second reader, the paper went to a third reader. If two of the scores matched, the paper received that score. If all three scores disagreee, the paper was discussed by the whole group.

c) In addition to the initial ten papers scored by the entire group, six packets of six papers were scored. The entire scoring session was tape recorded for later transcription. Because two of the scorers were forced by another school commitment to leave early, they were not present for the discussion of the last few papers. An analysis of the entire session then consists of three portions: the papers scored through group consensus, the papers scored using normal procedures with discrepant papers being discussed by all five scorers, the papers scored using normal procedures with discrepant papers being discussed by three of the five scorers.

d) Record was kept of the scores given by each rater, the match between raters, the outcome of the third readings and of any discussion. The scores were then compared to the scores given the papers in 1981 or 1982. Speculations concerning the differences in the scores were made by the participants.

e) After the tapes were transcribed, I examined the types of statements people made, the roles they played, and the strategies they used during the discussions. To make this examination, the types of statements made during the discussions were categorized, coded, and tabulated.

The Interviews

To explore in more depth the effect of an evaluation procedure on the teachers, I then interviewed the five participants in their schools. During the interview the participants were asked to do two things:

a) Score two papers, one that had received a one in its first scoring even though its surface features were practically flawless, and a paper that had received a four in its first scoring despite many surface errors.

b) Answer nine questions regarding the scoring procedure, the effects of primary trait scoring on themselves, and their perceptions of the effects that the scoring procedure has had in the school system. (See Appendix B.)

These interviews were tape recorded and transcribed. The methods and rationales used to score the two papers were compared across the five people and then compared to the methods used by the people when they were involved in the group scoring session. The answers to the nine questions were used to add perspective to what had happened during the scoring procedure as well as to gather information about how the evaluation process affects instruction.

Summary of Results - The Effects of Secondary Traits on the Scores

In order to analyze the effects of secondary traits on the scores that the papers received, three procedures were used. The first was the most direct: an examination of the sample papers. When 1981 and 1982 papers were examined, the findings were as follows:

1) Formula Beginning and Ending Sentences. Papers scored one, two, or three generally had topic and clincher sentences that followed the types of patterns recommended in students' English textbooks. The topics did become more specific with the higher scores. Papers scored the highest either had specific topics with the topic sentence/clincher sentence pattern or they had highly personal, engaging opening sentences and closing sentences that did not follow the textbook pattern.

2) Length. There is a difference between the 1981 and 1982 papers in the range of lengths in each score, especially in the three score. Overall, however, the tendency is for longer papers to be rated higher on the scale, even though papers of similar lengths appeared in all four scores for both years.

3) Spelling. There is a difference between the 1981 and 1982 papers when unconventional spellings are considered. In 1981, the highest average number of unconventional spellings and the widest range of the number of unconventional spellings for the ten sample papers fell in the three score. In 1982, the highest average number of unconventional spellings and the widest range appear in the one score. When the two years are combined, the tendency again is for the higher scored papers to contain fewer unconventional spellings. There were, however, unconventional spellings found in papers of all ratings. The 1981 and 1982 papers differed sharply in the types of unconventional spellings found in each rating, but for both years the majority of the unconventional spellings can be described as functional spellings: visual or phonetic approximations of the conventional spellings.

4) Punctuation and Usage. The papers rated one, two, and three have similar averages of number of types of errors per paper, while there is a decrease in the average number of types of errors in the four papers. The papers increase in average number of errors per paper from the one through the three rating with a drop in the average number of errors in the four papers. Again, virtually error-free papers are found in all scores, just as papers with errors are found in all scores.

5) Syntactic Complexity. The mean length of t-units, the mean length of dependent clauses, and the mean number of dependent clauses were virtually the same for all four scores. The mean number of t-units for the combined 1981 and 1982 sample papers did increase with the score, which was consistent with the findings concerning length of the papers: the longer papers tended to be in the higher scores.

When papers rated one, two, three or four from the two years are compared, there are similar papers in all four scores in terms of length, conventional spelling, command of

conventional punctuation and usage, and syntactic complexity as determined by t-units. The papers rated four do, on the average, show a greater command of the conventions and a tendency to be longer. There is a definite sense that unconventional features do not influence the raters as they consider papers in the lower scores, but those unconventional traits may influence the scorers to downgrade an effective paper from a four to a three.

The second procedure used to examine the effects of secondary traits was the taping, transcribing, and coding of a scoring session to see if secondary traits were a factor in group decisions. None of the participants introduced spelling, mechanics, or usage to convince others of a paper's score. There were times when unconventional spelling was obliquely mentioned, as when readers pronounced words according to the student's spelling. It was clearly an agreed upon rule of the scoring session, however, that secondary traits were not to be discussed to negatively affect a score.

The individual interviews were a third way to examine the effects of secondary traits on the scoring procedure. The differences in readers became very clear. Unconventional spellings interfered with the reading for some of the participants, causing them to question what the writers meant. Some of the participants admitted that they had to learn to disregard secondary traits when scoring the compositions; some thought that not all of the other scorers were able to disregard secondary traits while reading.

Issues of Reliability

The participants in the rescoring session were very anxious about the reliability of the procedure: would papers receive the same score the second time they were scored despite the intervening years? The readers have always viewed the scores as lying on a continuum, allowing some twos, for example, to be better than other twos. The participants believe that the difference in the readers may cause one person to rate a paper a high two while another rates it a low three. With this image of the scoring guide, the raters are not unduly upset by scores that differ by one number.

Using the participants' image of the scoring guide, the rescoring of the 1981 and 1982 papers shows this method to be reliable. Twenty-four of the forty-six papers (52%) that were rescored received the same score that they had received earlier. Twenty papers (43%) received either one score higher or one score lower than they had in 1981 or 1982. Only two papers (4%) received more than one score difference in this new session; both were rated more highly in the 1984 session than they had been scored in 1981. Although the rescored papers do not have high reliability in terms of an exact match with the previous scores, the raters were satisfied with the method's degree of reliability because of their image of the scores lying on a continuum.

The Effects of Individuals on the Scoring Process

After the scoring session was transcribed, I coded the remarks of each person during the session in order to examine how decisions were made concerning the papers under discussion and to chart how earlier discussions might have influenced later scoring decisions. The categories for the comments came from assigning a description or purpose to each statement. A sentence could be coded multiple times depending on the purposes that sentence fulfilled. Comments were coded according to these fourteen descriptors:

- A statement that a reader had a problem with the paper
- Placeholders (eg., simple statements of "I like this paper," or simple agreement with another scorer's comments)
- An analysis of the author's technique or style
- Agreement with another scorer's comments including an example from the paper
- Agreement with another scorer with additional reactions that lead to new avenues of discussion
- Surface agreement with exceptions ("yes, but...") or open disagreement
- Reference to the scoring guide to persuade, to solve a problem, or to defend a position
- Personal reactions to the paper or to a scorer's comments that may add new dimension or images to the reading of the paper (eg., "I feel that this kid could be a real clutz on

ice, but that he's definitely been ice skating, but that these things he describes on ice are those things that he wants to be on ice...")

- Quoting lines or examples in a paper to support a point
- A comment that is interrupted so that it is not followed up by any other participant, or a statement that is ignored by the other scorers
- A statement giving a score, encouraging a score, or finalizing the negotiation of a score
- A statement simply defending the author's belief or success
- A comparison to another paper
- A question for clarification

The types of statements with the most effect on the outcome of the discussions were those that analyzed the writer's technique or style, those that provided examples to support or expand points, those that made reference to the scoring guide, and those that gave personal reactions which added new dimensions or images to the papers. By using the coding system, I could chart the flow of the discussion, noting the strategies used by the readers who most influenced the outcome of the discussion, the variation in the strategies of each reader from paper to paper, the impact of earlier discussions on later ones. Since there were three distinct phases to the rescoring session, the coding system allowed me to examine the differences in the discussions during those phases. In addition to analyzing the number and types of persuasive statements made by each reader during the discussion of each paper, I also examined the turning points in the discussion of each paper to see the interplay between raters and any relationships that might be evident between the papers. A turning point was identified at a time in the discussion when the readers were deciding on a final score but refocused on an aspect of the paper that changed their views towards that final score.

From the analysis of the discussions in the scoring session, the following observations can be made:

1) In all but a few cases, the percentage of persuasive statements made to the total number of statements made by an individual is high (50% or better) for those raters who had their way in determining the final score.

2) Decisions tend to be made by majority vote. There is an attempt by those in the minority to persuade, but as time pressures mount, the majority simply outtalks the minority to settle on the score.

3) Individuals differ greatly in their strategies during the discussion. The participant who has the greatest impact on the final scores is the most consistent in use of persuasive statements throughout. A second participant changes strategies for each segment of the scoring session. A third alternates strategies from paper to paper, being persuasive on one paper, quiet on the next.

4) The three discussion segments during the single scoring session are distinct events. For example, when the raters work with the first group of papers (the group whose scores were determined by consensus), there is always at least one statement of a reader having trouble with the paper under discussion. Once the normal scoring procedures are in effect, there are no such statements of uncertainty. After two readers left early, creating in effect a new discussion group of three, the statements of uncertainty again arise. A second example of the shift in strategies with the change in discussion segments is evident in that no one in the new group of three ever analyzes the author's technique, perhaps because those people most swayed by that argument are no longer present.

5) The readers have a sense of orchestrating the negotiation process. The participants who most often persuade others to their view of the paper choose the type of statement, and the timing of those statements to lead the others to their way of thinking.

6) The readers operate with the sense that the papers are not isolated entities to be evaluated, but that they are part of a larger group of papers. Arguments made and information shared in discussion of earlier papers affects the discussion and ultimate scores of subsequent

papers. Those who are most skilled at influencing the scores selectively remind the others of those earlier papers.

7) The readers create new images and new visions of the papers by bringing personal experiences and attitudes towards the topic of the discussion. The most persuasive scorers are also the most adept at creating stories as a frame for the student compositions.

8) The scoring guide is a beginning point to the discussion and an authority to cite when a persuasive statement is needed. The scoring guide may temper and focus the reader's reactions, but it does not regiment those reactions. It is, in some cases, as subject to re-interpretation and embellishment as the student papers.

The Effects of the Scoring Process on the Individuals

During the interview, each participant was asked to think aloud while scoring two papers. In these individual interviews, the participants reacted to the students' papers in unique ways, interpreting those papers in accordance with their own perceptions of their roles as readers, teachers, evaluators, and participants in the group. The participants' comments as they read the papers highlight the interaction of those roles. Every participant expressed an awareness that involvement in primary trait scoring is not a one-sided process. All acknowledged that they, too, are affected by the process, changed by their interactions with the other scorers, and so, in essence, they have become learners. This awareness colors their reactions as individual readers, enabling them to accept changes in others' perceptions as well as in their own.

During the individual interviews, the participants were asked two questions to focus their attention on the effect that their involvement with primary trait scoring has had on them as individuals:

1) Do you read your student papers in a different way since you have been involved with primary trait scoring?
2) How has your experience with primary trait scoring affected your teaching of composition in your classroom?

The answers to these questions made it clear that participation in primary trait scoring has been important to each person as a teacher. It has given new ways to view student papers, new tools to use to discuss student papers, and new teaching methodology to use within the classroom. It is also obvious from the scope of the responses that the participants have not all been changed in the same ways. They did not all agree on the role of mechanics in instruction in the classroom or in the grading of compositions. It is clear, however, that they all perceive themselves as having changed, and as having transferred that change into their classroom practices. One participant saw the change as a change in her attitude toward student writers. Another saw the change as one of adding new teaching techniques and of ironing out the problems with those techniques. The third saw the experience as support for a particular model for instruction, one that allows the students more time for writing and more opportunity to read and react to their writing. The fourth uses primary trait scoring as an activity within his classroom, putting the students through the same experience that he values. The fifth emphasizes care in the wording of assignments and in the use of examples in teaching writing.

That all the participants in the primary trait scoring session perceive a change in themselves as readers of student compositions and as teachers of writing is clear from their responses. It is also clear that just as different readers take different things from what they have read based on their needs and their outlook, these individuals have learned different things from their involvement and have employed their new understandings in varying ways in their roles as classroom teachers and developers of curricula.

The Effects of the Scoring Process on the School District

The influence of this particular evaluation experience has gone beyond the scope of the reporting of a series of scores. The experience has had an impact on these evaluators' classes, and, because four of the readers have roles as administrators as well as teachers, it has had an impact on instruction throughout the district. Inservices for the English departments in each middle school have been held to acquaint other teachers with primary trait scoring.

Because the evaluation tool was in place in the middle schools, there has been more emphasis on writing instruction at the elementary levels, giving younger students writing experiences that would help prepare them for middle school. Individual teachers have used primary trait scoring in their classrooms to evaluate student compositions and to act as a springboard to a discussion about effective writing. There have even been inservices for administrators on how to write teacher evaluations using primary trait scoring as a teaching tool. Evaluation in this case is not an end checkpoint of curriculum and instruction, but a pervasive, on-going influence in the school system.

Primary Trait Scoring as a Transactive Process

Evaluators with Multiple Roles

In primary trait scoring the scorers take on roles in addition to their evaluator role. When the scoring guide was initially developed in 1978, many participants felt that time would be wasted in developing the scoring guide from descriptions of the papers since, as English teachers, they knew what would constitute an effective paper. The creation of the scoring guide forced the participants to change some of their English teacher expectations. Through the process of reading the papers, reacting to them, discussing them, and finally attempting to describe them, the scorers learned to trust their reactions as readers rather than relying on textbook-originated guidelines for good papers. As real readers they began bringing their own experiences to the reading. Before their involvement with primary trait scoring, the image of the teacher as reader was one of an objective judge listening to the evidence presented by the writer while waiting to be convinced of the defendent's merit.

Now the scorers are involved in the process, using their experiences to envision the writer's topic. The evaluation process freed the scorers to react as readers and to be honest in acknowledging the subjectivity of that reaction rather than clinging to a pseudo-objective stance of English teacher.

New experiences will necessarily mean new reactions as a reader. Even if the scorers remain constant from year to year, they are necessarily different readers because of the intervening experiences. The change in the raters as readers because of their involvement in the process goes beyond the model of primary trait scoring described by Lloyd-Jones (1977) who states that "Once the guide is created for a particular exercise...then presumably any bright and verbal person can use it." The creation and subsequent use of the scoring guide forces the readers to calibrate their reactions to the texts and to articulate their reasons for those reactions. As individuals, the readers confront the differences between their personal beliefs about writing competence and the evidence presented by the papers and by the positions of the other readers. While the individual reader may still value control of conventions in a student's first draft, for example, the scoring guide forces the reader to either ignore those private beliefs or search to find other evidence to support those personal reactions. That the raters are continuing to learn to read as individual readers even after six years of involvement with the process is evident from the transcript of the scoring session and from the individual interviews.

The participants may not react solely as readers; they do have the primary responsibility for giving each paper the score corresponding to the narrative that best describes that paper. Since two out of three scores must match, the scorers must keep personal reactions as readers within the agreed upon descriptions in the scoring guide. Because of this, the participants are also constantly renegotiating for themselves the meaning of the scoring guide as it relates to the individual papers. The redefinition of the scoring guide is also accomplished through the discussions that occur after papers' scores are not settled by third readers. The papers that go to discussion obviously highlight descriptions of three different scores, and the resulting discussions allow the participants to compare their understanding of the scoring guide to that of the other scorers and to refine their understanding according to the outcome of the discussion.

The papers discussed then become examples in later disagreements, adding to the meaning brought by the individuals to the scoring guide. As in the case of the participant's role as a reader, the

participant's role as a scorer is in constant change with new dimensions being added and with refinements taking place throughout the scoring procedure.

The interactions of the evaluators as rater are not addressed in Lloyd-Jones's description of primary trait scoring. It is clear from the transcript that within the group, the raters use negotiation strategies to sway the other readers. Primary trait scoring is not only dependent on the raters as readers, but on the mesh of their personalities, their negotiating skills, and the social interactions among the group. The constant re-visioning of the scoring guide in relation to the papers scored demonstrates that knowledge is socially constituted. The scores given papers within a single session represent not simply the written description on the scoring guide, but the scorers' negotiations and adjustments to each other's interpretation of those descriptions.

The participants act not only as readers and scorers, but also as learners. The interviews made clear that all five realized a change in their beliefs about writing and about evaluation because of their involvement in the process. The interviews also clearly show that not all the participants have learned the same lessons or have acquired the same set of beliefs about writing or have the same understanding to be able to talk about what they have learned. Involvement in the process has prompted a re-evaluation of beliefs on the part of all the scorers, but it has not provided for a uniform change in beliefs.

The participants also bring with them their role as instructors. Their experiences with primary trait scoring have provided new ways of explaining to students why one paper is more effective than another without relying only on criteria that can be quantified such as spelling or usage errors. The negotiation session becomes a place to try new explanations and to refine definitions and examples of effective writing. During the session, concerns with how to explain particular decisions about scores arose. During the interviews, participants clearly were aware of the effect their involvement has had on them as teachers.

The transfer of lessons learned through involvement in primary trait scoring to classroom instruction is tempered by concerns for the perceived values of colleagues, principals, and parents.

A sense of tension is created between, on one side, the ideas for writing instruction springing from the new understandings about writing and, from the other, a sense of responsibility for upholding these perceived community values of correctness and convention. The teachers involved as raters ease the tension in different ways. Some have the students become involved in the scoring process so that they experience the same lessons as their teachers. Others use the primary trait scoring experience as justification for traditional teaching methods. While the participants come to new understandings about what constitutes effective writing, as teachers they are left in a void; they find themselves alone in translating the new understanding into teaching practice. Because the impact on instruction by an evaluation practice is not formally recognized and explored, teachers are left to make curricular decisions without the support of a uniform theoretical model or philosophy. The impact of the evaluation procedure on classroom instruction, then, is not uniform across raters. Instead, it varies according to the perceptions and constraints felt by the raters as individual teachers.

Since readers bring different experiences as readers, different understandings of the scoring guide as raters, different beliefs as learners, and a sense of responsibility to their roles in varying teaching and administrative roles, the composition of the group is critical to the outcome of the session not so much in terms of the scores the papers receive, but in terms of further redefinitions and changes in the participants' roles. The papers may be a constant, but the evaluation of them is not. Primary trait scoring is not a static procedure causing no change in either product or participant. Primary trait scoring as used by this school system is a dynamic process causing far reaching changes in its participants and in their classrooms.

The Social Context of Evaluation

Even if the same people are involved each year as scorers, a social context is established with each session that influences the flow of the discussions. Three of the participants are English department chairpersons who have attended many other meetings together with the fourth participant, the district's curriculum coordinator. These people will all have to work with each

other again in the future. Reluctance to negotiate scores or actions that could be perceived as a stubborn refusal to accept another's viewpoint could damage working relationships needed for other meetings, other settings.

The discussions of the papers reflect not only the participants' varying roles, perspectives, and strengths, but also an awareness of what the other participants bring with them to the scoring session. The attempts to persuade others of the effectiveness of a paper are tailored for the particular set of participants and for the necessity to establish and keep good working relations among the group by mutual compromise. When a paper has received three different scores and comes before the group for discussion, it is not read in isolation. That paper is discussed in relation to what has been discussed and decided before and according to the time pressures sensed by the group within the social setting. The result of the discussion then is the creation of a new, agreed upon understanding or "poem" to use Rosenblatt's term. This new "poem" is based on the student's text, colored by the perceptions of the individual readers, imbued with the images created by the scorers to persuade one another, suffering or benefiting from the negotiations that have preceded it and curtailed by the amount of time left for discussion. All of this occurs within the social setting of the scoring session which itself occurs within the broader context of the curricular changes that have and will take place because of this evaluation procedure.

Summary

Primary trait scoring, as used with this assignment and these scorers, capitalizes on the strengths of the participants as readers, scorers, learners, and instructors within the social setting established through a knowledge of one another and of their common sense of purpose. The evaluation process causes the participants to rethink their beliefs about writing and writing instruction, and more importantly, to continue to rethink, examine and refine those beliefs through the interaction of the group discussions. Because the method validates the use of whole texts and because the participants act as leaders for change within their school, more teachers are involved in having students write whole compositions, and curricular and methodological changes occur in the

classroom. This study, which examines the papers from only two years, cannot address the question of how the changes in the classroom have and will affect the quality of student papers. It is clear, however, that this evaluation technique has a direct effect on the teachers involved and on the instructional methods used in classrooms throughout the school district.

REFERENCES

Harste, J. and Burke, C. (1977). A new hypothesis for reading teacher research: Both the teaching and learning of reading is theoretically based In P. David Pearson (ed.), Reading: Theory, Research and Practice (26th Yearbook of the National Reading Conference). Rochester, NY: National Reading Conference, 32-40.

Harste, J., Woodward, V., Burke, C. (1984). Examining our assumptions: A transactional view of literacy and learning. Research in the Teaching of English, 18, 84-108.

Lloyd-Jones, R. (1977). Primary trait scoring. In Charles Cooper and Lee Odell (eds):,Evaluating Writing. Urban, IL: National Council of Teachers of English.

Rosenblatt, L.M. (1978). The reader, the text, the poem: The transactional theory of the literary work. Carbondale: Sourthern Illinois University Press.

APPENDIX A

THE SCORING GUIDE

The reader should note that the following scoring guide refers to a paper rated zero. This rating is not a quality judgment; the papers are not necessarily better or worse than any other score. The zero category is a non-judgmental rank given to a paper that cannot be scored according to the scoring guide because the writer did not write on the assigned topic. The zero rating is not considered part of the continuum of scores from one to four that raters use to denote the writer's effectiveness. Since these papers were considered by the readers to be unrated, they were not included in this study.

Score Description

0 These papers avoid the assignment. The writer does not establish an attitude. Paper 134 ("Saxaphones") is an example of a "0" paper.

1 These papers respond to the assignment. The writer's attitude is established (generally, an "I like..." or "I dislike..." statement is made.) The writer's attitude is not developed or a confused attitude is expressed. Details or description do not relate to the writer's purpose. There is little personal involvement expressed. These papers do not evoke an emotional response in the reader. Papers 217 ("Brautworst") and 18 (The Hobbit") are examples of "1" papers.

2 These papers respond to the assignment. The writer's attitude is established. "I like" does not constitute establishment of attitude. Specific description or details that relate to the writer's purpose are included. These papers lack elaboration. Repetition of the same details or description and/or randomness may detract from the effectiveness of these papers. There is an attempt to evoke an emotional response in the reader. Papers 55 ("Snapping Turtle"), 156 ("Go-Kart"), and 39 ("Macaroni and Cheese") are examples of "2" papers.

3 These papers respond to the assignment. The writer's attitude is established and the attitude is developed through elaboration. Elaboration may be specific description or details that focus on the object, use of metaphor, a narrative about the object, or strong sensory appeal. The reader feels the writer's intense personal involvement. These papers evoke an emotional response in the reader. The "3" paper lacks the unity and style of "4" papers. Papers 128 ("My Fawn"), 21 ("Knuckles"), and 23 ("Spinach") are examples of "3" papers.

4 These papers exhibit all the characteristics of the "3" papers. The writer's attitude is established and consistently and constantly reinforced. The "4" papers are developed in one of the following ways:

The writer uses a narrative to establish and maintain an attitude toward the thing being described.

The writer gives a vivid and powerful reason for his attachment to the thing being described.

The writer uses specific, vivid description, and/or strong sensory appeal, and unusual or unexpected use of language to

communicate personal attachment or involvement with the thing being described.

These papers have unity. All details and description contribute to the desired effect. There is no sense of randomness. The papers are tightly constructed and hang together. The "4" papers are cohesive. Sentences flow and there is a tight interrelationship between sentences. The style contributes to the overall effect; sentence structure and word choice are highly important to the subject matter. A strong emotional response is evoked in the reader. Papers 113 ("Desks") and 1 ("Gnats") are examples of "4" papers.

135

Example O

The saxophone is a very versitile intrument While being a very good solo intrument it is also an intrumeat that can add to some other intruments song. A talented saxophonist can make it sound harsh or mellow.

The saxophone was invented by Antoine J. Sax. It is a new intrument that is still being chaned today For instance a low B flat key was added very recently. There are also new fingerings being added to make the two higher B flats.

A saxophone is played by blowing air through a mouthpiece. There is a reed attached to the mouthpiece. The reed vibrates and seeds a sound through the brass body. The sound is altered by pressing keys which open and close holes in the body. Although called a woodwind intrument the saxophone was never made of wood.

There are several kinds of saxophones. There is a soprano sax (short for saxophone), an alto sax, a tenor sax, a baritone sax, and a bass sax. I play the saxophone and like it very much.

18 Example 1

I like my fully illustrated edition of "The Hobbit." Inside is a beautifully told story of the adventures of a courageous hobbit. The story is filled with enough adjectives to last anyone a lifetime. When reading this novel at any time you may feel that you are experiencing an action packed scene. Once the hero gets in trouble it's almost impossible to stop reading until he's out of danger. On each page is a illustration or two showing the wonderful world of Middle Earth where this grand adventure is taking place. Each drawing is beautifully illustrated and filled with every color imaginable. It also has a carefully planned index so it is easy to find anything you wish in the story. When you have finished the book you will feel that it had a broader meaning. And also an example should all follow.

217 Example 1

English Composition

I hate brautwurst. The last five letters tells what it is = the worst. Even the word I hate. Say brautwurst for a long time and you'll realize how stupid it sounds. Furthermore, I'm not even sure I spelled it right. It's one of those toughies. Have you ever wondered what is in brautwurst. You may look at the ingrediants, and it will say that it's chuck full of nutreints and vitamins. Huh, I heard a story that the place where they churn all stuff in a big vat, a couple of men was stirring it up, one man went to answer the phone, when he came back the other man was gone. They never found him. Then a couple of weeks after that, some woman complained that one day she was eating a brautwurst cracker, a bone suddly appeared. This may not be true but I'd be awful carefull.

18

Example 2

Go Kart

Karting is a fascinating sport that the whole family can participate in. From little brother to cheer you on, up to mom and dad to help you in times of trouble.

I have a kart of my own. I had ~~to~~ worked in my uncles very own go-kart ~~sh~~ shop until I recived enough credit to get the cart. It is a kart that you would call an enduro kart. It has two Briggs and Stratton five horsepower engines, two three gallon rectangular ~~gasolene~~ gasoline tanks, which burn exotic fuel mixtures, twin exhaust system, and a fully padded reclining seat with headrest. Burning exotic fuel mixtures will allow your kart to be able to go one hundred miles per hour! Going at those speeds calls for safety equipment. You must wear a full cover helmet with a ~~bad~~ bubble attatched in some fashion to the helmet, a heavy leather jacket, durable jeans, elbow and knee pads, and last but not least, heavy boots.

I have a Mc Culloch chain saw

with a V belt attatched to the cutting head for a starter since you need a high rate of revolutions per minute to get the engine started.

I think that everyone should get into the sport of karting. Not only for the enjoyment, but you can also race at Indianapolis' Raceway Park.

Example 2

I dislike my brothers snapping turtle. It is still quit young, but he tries to bite. He also has very sharp long claws. When you try to pick him up he tries to scratch and when he does, the scratch is deep and hurts for a long while. His shell is hard and ruff.

39 Example 2

I hate macaroni and cheese. It looks terrible and tastes terrible I don't like the smell of it. It looks like something you'd pull out of a trash can. It looks very messy

23 Example 3

"Spinach"

I hate spinach so much it's hard to believe. I used to like it when I was little, but that was because Popeye ate it! I dislike it so much because of the horrible terrifying things about it.

#1: it has a sickening aroma, or I should say odor. When my mother cooks it I get shivers up my spine, and then I gulp it down so it doesn't ruin the rest of the meal

#2: spinach has a stomach churning taste. I don't know why, but it's just horrible and seems like rotten leaves that are soggy, with a little grass mixed in.

#3: spinach has an appearance that is very gross. That may be why I hate it. It looks like the inside of a squashed worm that needs to be eaten. yech!

That's what I think of spinach, the guy sitting across from me loves it. yuck !!!!

Example 3

One thing that I hate most is when someone cracks their knuckles. The noise distracts me terribly. It sounds like bones being broken in half or high heels clicking noisily down a hall. When someone cracks their knuckles all the tension builds up inside of me then look out, because I'll explode. Also I don't see why they do it. To get arthritis? Or to get big knuckles? I think it is the worst habit, next to picking your nose, that anyone could ever have.

12

Example 3

The object I am going to describe is soft, fuzzy, and very cute. It is about 1 foot tall. It has small white spots on its back. It is tawny in color. Another nice thing about it is it's long slender legs. Its rather a peaceful object. You may of seen them in the forest but mine is quite different. It is the most beautiful object in my house. It is very distinguished looking. The object I am talking about is a stuffed fawn. And its almost my favorite thing in the whole world. It is as if a fairy brought it to me to bring me happiness. It doesn't have any flaws. My fawn is a doe. My doe is fabulous.

Example 4

May 29, 1979

"Gnats"

Some things that I distike the most are gnats. They buzz all around you with a high, soft buzz and they dive bomb over your head so that you have to swat them away. Also, since I wear glasses, they land on my glasses and I see the black spot in front of my eye. So I try to make it fly away by putting my finger in my glasses but I end up smashing it so I have to wipe it off and all it does is smear all over.

They also land in your ears. Oh! what a pain!!! They go someplace where your finger can't get it but it tickles like crazy. So you have to sit there laughing to death until it decides to come out. They are enough to drive you INSANE!

Example 4

113

"When was the last time you've tried to sit on one of the school's desks for an hour, or for that matter seven hours?" That's what I'd love to ask the principal. Why? Because I don't think he (or she) would enjoy trying it again.

First of all they're made of probably the hardest wood they could [illegible], so it's not [illegible] comfortable. Secondly the backs are made (or so it seems) to cause you pain. You sit up and it pokes you, you slouch down and you get "zapped" by a teacher. Any way you look at it it's a losing proposition. Thirdly, if you've ever put your fingers under a desk what do you find? That's right, three year old gum. But if you don't look at the gum it

feels more like little rock
glued to the bottom of the
desk! Lastly, and maby worst
of all, is a thick metal bar
that leads from the right
side of the ~~of~~ bottom of the
desk, to the right side of
the ~~mac~~ small iron pedistal.
This permits you to bump your
knee as much as any
masocist would want to.
It also retards any ~~bac~~ leaning
back, I understand why
they wouldn't want you to
lean back but who cares?

APPENDIX B

Interview questions concerning scoring procedures:

1. What features in a student's paper do you usually notice first when you score a paper?

2. Do certain topics affect you as a reader? Which ones? Why? How do they affect your scoring?

3. What seems to cause discrepancies between the scores you give a paper and the scores other readers give that paper?

4. From your perspective, how has the scoring changed for these papers over the years?

Interview questions concerning the effects of primary trait scoring on the reader/scorer:

5. Do you read students' papers in a different way since you have been involved in primary trait scoring?

6. How has your experience with primary trait scoring affected how you teach composition in your classroom?

7. Looking back at our initial experience with primary traiting, what do you remember about changes in the participants' viewpoints as the week progressed?

Interview questions concerning the effects of primary trait scoring on programs/curriculum/methodology in MSD Wasington Township:

8. In your opinion, what impact has primary trait scoring had on Washington Township?

9. What are your reservations about primary trait scoring?

Is there anything else you would like to add?